ACADEMIC WORDS IN ABSTRACT OF UNDERGRADUATE THESIS

Rezza Dewintha[1], Sudarsono[2], Ikhsanuddin[3]
Masters Study Program of English Language Education
Teacher Training and Education Faculty, Tanjungpura University, Pontianak
rezzadewintha@yahoo.com

Abstract

Corpus linguistics is the study of language data on a large scale, computerization data analysis of extensive collections of written texts. This study is a academic words in abstract of undergraduate thesis of English Language Education Study Program, Teacher Training Facutly, Tanjungpura University, submitted in year 2012-2016, using AntConc Program. There are six (6) research questions raise in this study; (1) the categories of POS tagset; (2) the most dominant POS tagset; (3) The most dominant academic word; (4) the role of AWL in students' abstract; (5) the most dominant AWL sublist; and (6) the comparison between AWL and non-AWL. The findings are; (1) there are 31 types of Tree Tagger tagset and Brown Corpus Tagset; 11710 tokens (18,2%) for Noun, 8198 tokens (12,8%) for Determiners, 5289 tokens (8,2%) for Preposition, 4075 tokens (7,3%) for Verb, 2410 tokens (3,7%) for Adjective, 809 tokens (1,3%) for Conjunction, 527 tokens (0,85%) for adverb, 2 tokens (0,003%) for Interjection; (2) NN as the most frequent tagset, (3) the most frequent academic word is Research, (4) only 0,00522% or 183 types of academic words used by undergraduate students, (5) sublist 1 as the most frequent sublist used, and (6) there are 1974 non-academic words used. There are 208 corpora which are gathered from UP4I since 2012-2016.This research is useful for English Language teacher in examine their students' writing easily, especially focus on concordance entirely and lexical study. It leads to more principled classroom materials and activities and lead students to have best understanding on language analysis.

Keywords: Corpus linguistics, Lexical study, AWL

Corpus-based approach to linguistic analysis and language teaching has come to prominence over the past two decades. By developments in technology, especially the development of over more powerful computers offering ever increasing processing power and massive storage at relatively low cost, the exploitation of massive corpora became possible. Corpus-based research assumes that validity of linguistics forms and structures is derived from linguistic theory, and the primary goal of research is to analyse the systematic patterns of variation and use for those pre-defined linguistic features (Biber, 2012, p.1). Then corpus is a collection of (1) machine-readable; (2) authentic texts (including transcripts of spoken data) which are (3) sampled to be (4) representative of a particular language or language variety (McEnery, Xiao, &Tono, 2006, p.5). Therefore, corpus linguistics is the study of language by means of naturally occurring language samples; taken from real word context. Their analysis is usually carried out with specialized software program on a computer; in this research, the researcher is using AntConc program from Laurance Anthony to analyze the using of Academic Word List (AWL).

Corpus linguistics is invaluable for non-native teachers, because the great advantage of the corpus linguistic method is that language researchers do not have to rely on their own or other native speakers' intuition. Rather, they can draw on a large amount of authentic, naturally occurring data produced by a variety of speakers or writers. Moreover, the essential qualities of corpus include machine-readability, authenticity and representativeness (Aston & Burnard, 1998, p.5).

Hence, the significances of conducting this research will emerge of course in academic field. The researcher realizes that being university students will face a lot of obstacles during writing scientific paper. There are demands as a student; he/she needs to submit paper in every single subject as an assignment, and he/she needs to submit an article into national or international conference. The students did not know how to elaborate the academic words into their words because of the lack of academic vocabularies. With students' deficiency of academic word list, they could not advance their writing. To overcome that inadequacy, the researcher thinks that is interesting to conduct this research. Researcher wants to investigate how the students use academic word lists in their paper, especially in an abstract of journal article. Also, the researcher wants to investigate the word frequency and the distinction of the use of academic word and non-academic word in their abstract journal article.

The researcher wills mentions several advantages relevant to this research. For instance, students may enrich their lexical knowledge regarding Coxhead's Academic Word List (AWL) whether it is very important in academic life. They also know the AntConc program to analyse the corpus. Tognini-Bonelli (2001) found that "The methodology of corpus linguistic in classroom context is a "bottom up" study of language learning for all levels students; even for students with lack of linguistic knowledge, they quickly to advance their knowledge" (p.16).

With those great advantages which corpus-based research has, the researcher intends to apply this method in this research to analyze the use of Academic Word List (AWL) created by Dr. Averil Coxhead. The researcher would like to investigate; (1) the pattern of the Part of speech (POS) tagset; (2) the most dominant POS tagset appears during this research; (3) the most dominant academic word used by students; (4) the most dominant AWL Sublist appears during this research; (5) the most dominant academic word appears during this research and (6) the amount of non-academic word used by undergraduate students. This will be investigated due to their importance in academic writing. On the basis of all university students must have good writing skill and also writing products. Then in the future, the researcher will know the undergraduate students' ability in using Academic Word List (AWL), and also the pattern of part of speech. Hence, the researcher want to conduct a research entitled, "Corpus-based Lexical Study on Academic Word List (AWL)" (A Corpus-Based Lexical Study Using AntConc Program on Undergraduate Students' Abstract Journal Article of English Language Education Study Program, Teacher Training Faculty, Tanjungpura University, West Kalimantan, Indonesia, Submitted in Academic Year 2012-2016).

METHODS

This research is designed with quantitative corpus linguistics to answer the research questions listed above. In this research, the corpus establishment were built and acts as a database; a written specilized corpus containing 570 words families, 3.500.000 words of text and those are divided into 10 sublists running words from 208 written texts of a single genre – English Language Education research especially undergraduate abstract journal article.

As discussed in the beginning corpus linguistics which is primarily defined as its preferred method; the dominant data source is naturalistic spoken and written data (as presented in corpora), and accordingly the major methodological tool is corpus analysis (Krug & Schluter, 2013, p.10). In the corpus linguistics, there are two types of method; corpus-driven and corpus-based. After proof reading, the appropriate method for this study is corpus-based study. Corpus-based study that have examine the linguistic means of information highlighting in English interlanguage perspective (Dobric, Graf, & Onysko, 2016, pp.15-16). This study aims to investigate the ability of L2 learners in using AWL in their abstract journal article. The researcher would like

to investigate: the pattern of POS tagset that will be formed; the most dominant POS tagset and academic word appear during this research; the way students use AWL in their abstract journal article; the most dominant AWL sublist used by them; the distinction between undergraduate and postgraduate students' ability in using AWL.

FINDINGS AND DISCUSSION

Findings

The data derived from the 208 students' abstract journal article. The infromation about the frequent word in each tagset also will be enclosed. Realated with the part of speech data tagset, the result will be shown in table with clear explanation supported.The data result which shown the tokens of the word frequency will be devided into three categories (1) the most frequent word; (2) the frequent word (fx >100); and (3) The rare word (fx >50). The most frequent word means; the word stands in first rank in each tagset. Then the frequent word means word which has frequencies above one hundred (100) and the rare words means; word which has frequencies above fifty (50) but less than one hundred (100) tokens.

After analyzing the datam there are fourty five types of Brown Tagset appeared in this research. But only thirty two types of tagset have the tokens. These types are commonly encountered in students' abstract journal article. ths is provide an explanation of students' writing have multifarious tagsets. It is 51,5% used by students in their abstarct journal article. the rest is 48,5% stands for another tagsets. Furthermore, the answers of research wuestions can be seen in the subheadings.

How will the patterns of Part of Speech tagset will be formed?

Noun

As can be seen in the table 1. Table of Nounbelow, there are four types of tagset turn up in this research. The higest frequency is *NN* with 7784 tokens. The second is *NP* with 1996 tokens. The third is *NNS* with 1927 tokens, and the last is *NPS* with 3 tokens. The part of speech of *Noun* became the most frequent tagset appeared in students' abstract journal article. With 18,23 percentage in used.

Table 1. Table of Noun

Tagset	Description	Tokens	Whole Frequency
NN	Noun, Singular, Common	7784	18,23 %
NNS	Noun, Plural, Common	1927	
NP	Noun, Singular, Proper	1996	
NPS	Noun, Plural, Proper	3	

Based on *NN* tagset there are five hundred and seventhy (570) word types or word tokens. The most frequent word is *Research*came from Avrile Coxhead's Academic word list Sublist 1 with 748 tokens in students' abstract journal article.

There are nineteen (19) frequent words based on *NN* tagset. According to Avrile Coxhead's Academic Word List, there are six (6) academic words. Those are $data_{235}$ *and* $method_{100}$ from Sublist 1; $text_{230}$ from Sublist 2; $technique_{156}$ from Sublist 3; $cycle_{154}$ from Sublist 4, and $grade_{154}$ from Sublist 7.

The second from the part of speech tagset is *NP* with 430 word tokens (see Appendix VII, Table of *NP* Word List, page 272). The most frequent word is *Pontianak* with 104 tokens in students' abstract journal article. The detail can be seen in Table 2.

Table 2 Table of *NP* Frequent Word

Total Words Token	The Most Frequent Word	fx >100 (frequent word)	fx <50 (rare word)	
430	Pontianak (104)		English	71
			Negeri	52

There are seven (7) frequent words based on *NP* tagset. There is no academic word in the frequent words used by students. The $Pontianak_{104}$ there is proper noun. Proper noun belongs to noun that its primary application refers to a unique

entitiy, such as naming people, animals, places, things, and ideas. Because of the corpus gathered from students who studied in Tanjungpura University especially in Undergraduate Program of English Language Education, teacher Training and Education Faculty which is placed in Pontianak, therefore the mostly naming place mentioned is *Pontianak*.

The third from the part of speech tagset is *NNS* with 198 word tokens (see Appendix VIII, Table of *NNS* Word List, page 277). The most frequent word is *Students* with 600 tokens in students' abstract journal article. The frequent word is $data_{235}$ belongs to Sublist 1 of Coxhead's academic Wordlist. The detail can be seen in Table 3.

Table 3 Table of *NNS* Frequent Word

Total Words Token	The Most Frequent Word	fx >100 (frequent word)		fx <50 (rare word)	
198	Students (600)	data	235	are	97

The last form of *Noun* tagset is *NPS*. It has two (2) tokens. The first one is $Sciences_2$ and $Parkinsons_{1.}$ To conclude, the whole percentage for *Noun* tagset is 18,23% stands in the first rank in this research as the most frequent tagset appear in students' abstract journal article.

Determiner

The second rank is *determiner* tagset. There are four (4) types of tagset underpin. They are *DT, CD, PP$,* and *AT*. The detail will be shown in table 4.

Table 4 Table of Determiner

Tagset	Description	Tokens	Whole Frequency
DT	Determiner, Singular	3640	12,8%
AT	Article	3004	
CD	Numeral, Cardinal	1470	
PP$	Determiner, Possessive	84	

From table 4, there are significance used by looking at the amount of *determiner* tagset appeared in students' abstract journal article. The whole percentage is 12,8% with total tokens is 8198. The explanation will go throughout the determiner tagset. The first will be *DT* tagset. The following detail will be shown in the table 5.

Table 5 Table of *DT* Frequent Word

Total Words Token	The Most Frequent Word	fx >100 (frequent word)		fx <50 (rare word)
7	the (2724)	this	571	
		a	209	

Determiner are words or signs that modify noun; for example *the, a,* and *an* are determiners in English (Valli & Lucas, 2000, p.103). Determiner plays an important role in writing epecially in academic writing. Because determiner as a grammatical relation while maintaining a distinction between heads and specifiers (Ghomeshi, Paul, & Wiltschko, 2009, p.5). Therefore, the using of determiner is high in students' abtract journal article.

The second tagset after *DT* is *AT* tagset stands for *article*. In English, there are two types of article. They are definite and indefinite article. Hence, the result for *AT* tagset is the_{2734} as the definite article, and a_{227} as the indefinite article. The third tagset in *determiner* is *CD* stand for numeral and cardinal numbers. Because there are a lot of types of *CD* appeared in students' abstract journal article, the result will be attached in Appendix X Table of *CD* Concordances in page 298 The *CD* tagset has 1470 tokens with variety of numerals and cardinals.

Preposition

The third rank in part of speech tagset is *IN* means preposition. It has 5289 tokens and the whole frequency is 8,2% used by students in abstract journal article. Moreover, there are thirty (30) types of preposition appeared in this tagset. The detail can be shown in the table 6.

Table 6 Table of Preposition Word List

No.	*IN* Words List	Tokens
1	of	1499
2	in	1012

3	by	242
4	on	242
5	as	228
6	with	120
7	for	107
8	from	99
9	through	88
10	at	62
11	about	34
12	after	27
13	between	26
14	because	16
15	into	15
16	during	13
17	before	12
18	Toward	10
19	Up	8
20	among	4
21	below	3
22	Next	3
23	Until	3
24	above	2
25	although	2
26	Within	2
27	despite	1
28	if	1
29	to	1
30	Since	1

From this table the result can be classified based on the most frequent word, frequent word (based on fx >100) and rare word (based on fx < 50). The detail can be seen in the table 7.

Table 7 Table of *IN* Frequent Word List

Total Words Token	The Most Frequent Word	fx >100 (frequent word)		fx <50 (rare word)	
30	of (1499)	in	1012	from	9
		by	242		
		on	242	through	8
		as	228		
		with	120		
		for	107		2

The preposition *"of"* is the most frequent word used by students. Because *"of"* can show the concerning about somethingand also connected with indicating origin and source while writing.

Verb

The fifth rank in part of speech tagset is *adjective*. It has 4075 tokens and the whole frequency is 6,3% used by students in abstract journal article. In this part, there are fifteen (15) tagsets appeared in students' abstract journal article. They are, *VB, VBD, VBG, VVN, VVG, VBZ, VVZ, VVP, VBP, VHD, VHP, VHZ, VBN, VH,* and *VHG*. The detail can be seen in the table 8.

Table 8 Table of Verb

Tagset	Description	Word Tokens	Whole Frequency
VB	verb *be*, base form	77	6,3%
VBD	verb *be,* past form	846	
VBG	verb *be,* gerund/participle	4	
VVN	verb, past participle	941	
VVG	verb, gerund/participle	1084	
VBZ	verb be, pres, 3rd p. Sing	588	
VVZ	verb, present, 3d p.sing	219	
VVP	verb, present, non-3rd p.	116	
VBP	verb *be*, pres non-3rd p.	95	
VHD	verb *have*, past	35	
VHP	verb *have,* pres non-3rd per.have	23	
VHZ	verb *have*, pres 3rd per.sing	20	
VBN	verb *be,* past participle	13	
VH	verb *have,* base form	7	
VHG	verb *have,* gerund/participle	7	

The information related to the word type, word token and also the frequent

word were analyzed very carefully and deeply.

Table 9 Table of The Most Frequent Verb Tagset

Rank	Tagset	Freq
1	VVG	1084
2	VVN	941
3	VBD	846

The variety of verb tagset in this case shown a dinamic used of the *verb* types in students' abstract journal article. From the table 9, the data shown *VVG* tagset is become the most frequent tagset used by students in their abstract journal article. With 126 word types and 1084 word tokens. The *VVG* tagset means *verb, gerund/participle.* Then, there are two words as frequent word (fx >100), they are *speaking*$_{111}$ and *teaching*$_{111}$. Moreover, the word *writing*$_{88}$, *learning*$_{72}$, and *reading*$_{71}$ as the rare words (fx >50).

The second rank in verb tagset is *VVN*. The *VVN* tagset means *verb, past participle.* It has 138 word types and 941 word tokens. The most frequent word is *used*$_{97}$. Then the rare word (fx >50) are *based*$_{92}$, *conducted*$_{71}$, *categorized*$_{61}$, and *collected*$_{58}$.

The third rank is verb tagset is *VBD*. The *VBD* tagset means *verb be, past form*. It has two word types, and 846 word tokens. The words are *was*$_{634}$ as the most frequent word used by students. Then *were*$_{212}$ belongs to frequent word with frequency above one hundred.

The fourth rank in part of speech tagset is *adjective*. It has 2410 tokens and the whole frequency is 3,8% used by students in abstract journal article. Moreover, there are three tagset underpin in adjective. They are *JJ, JJR,* and *JJS.* There are 313 word types regarding the adjetcive in students' abtract journal article. the following detail about the table of adjective and the table of *JJ* frequent word.

For the second and the third tagset of adjective; *JJR* and *JJS*. The *JJR* means adjective comperative. It has ten words token, and the frequent word is *higher*$_{44}$. Whereas the *JJS* tagset meand adjective superlative. It has four words token, and the frequent word is *most*$_{20}$.

The fourth rank in part of speech tagset is *adjective*. It has 2410 tokens and the whole frequency is 3,8% used by students in abstract journal article. Moreover, there are three tagset underpin in adjective. They are *JJ, JJR,* and *JJS.* There are 313 word types regarding the adjetcive in students' abtract journal article.

For the second and the third tagset of adjective; *JJR* and *JJS*. The *JJR* means adjective comperative. It has ten words token, and the frequent word is *higher*$_{44}$. Whereas the *JJS* tagset meand adjective superlative. It has four words token, and the frequent word is *most*$_{20}$.

The sixth rank in part of speech tagset is *conjunction*. It has 809 tokens and the whole frequency is 1,3% used by students in abstract journal article. There are four word tokens for this part of speech tagset. And the most frequent word is *and*$_{749}$ and the rare word is *or*$_{53}$.

The mostly used conjunction is coordinate conjunction which consists of *for, and, nor, but, or, yet, so*. But in this case, *and, or* and *but* became the most frequent used by students.

Adverb

The seventh rank in part of speech tagset is *adverb.* It has 527 tokens and the whole frequency is 0,82% used by students in abstract journal article. There are three types of tagset underpin this part of speech tagset. They are *RB, RP,* and *RBR*.

The *RB* tagset has eighty seven (87) words type, and 412 tokens. The most frequent word used by students is *not*46.

From this table, the data shown that there are various type of adverb used by students in their abstract journal article. The list consist of adverb with *–ly* and adverb without *–ly*. Based on the table, there are also comjunctive adverb; which connects two sentences and provides adverbial

emphasis. For instance, *therefore, then, rather* and so forth.

While *RP* tagset has two words type. Then the most frequent word appeared in students' abstarct journal article is *out*$_{83}$.

Out become the most freqeunt adverb particle appeared in students' abstract article. Again, there are nine (9) types of *"out"* clusters words. As shown in the table 10. The combination of *verb + adverb particle* often indicating direction, someties idiomatic.

Table 10 Table of *"out"* Cluster

No.	*"out"* Cluster	Tokens
1	find out	59
2	finding out	9
3	carried out	5
4	found out	2
5	try out	1
6	map out	1
7	trying out	1
8	tried out	1
9	finds out	1

The second adverb particle which appeared in students' abstract journal article is "up". The *"up"* often cluster with *sum*$_2$, and the rest cluster with *speak, catch,* and *going*. The detail can be seen in the table 11.

Table 11 Table of *"up"* Cluster

No.	*"up"* Cluster	Tokens
1	sum up	1
2	speak up	2
3	catch up	1
4	going up	1
5	increased up	1

Furthermore, *RBR*stands for the comparative adverb. The *RBR* tagset has five words type, and the most frequent word is *more*$_{16}$.

Interjection

The part of speech *interjection* is infrequently emerge in students' abstract journal article. In this research, there are only two tokens related to the *interjection*tagset. The tagsest for interjection is *UH*. The word is *yes*$_{2.}$

What does the most dominant POS Tagset appear during this research?

The most dominant POS Tagset appears during this research, is NN (noun) with 7784 tokens or 18,2% as the following tagset of part of speech.Then, as the most frequent noun in the first rank is*IN* with 6894 tokens.

What does the most dominant academic word appear during this research?

The most dominant academic word appear during this research is *Research* with 748 tokens comes from the first sublist. The second rank is *Text* with 230 tokens from second sublist. The third rank is *Technique* with 156 tokens from third sublist. The fourth rank comes from fourth dan seventh sublist, they are *Cycle* and *Grade* with 154 tokens. The fifth rank is *Found* with 40 tokens comes from nineth sublist. The sixth rank is *Interval* with 16 tokens comes from sixth sublist. The seventh rank is *Drama* with 11 tokens comes from eigth sublist. The eigth rank is *Equivalent* with 9 tokens comes from fifth sublist, and the last is *Odd* with 2 tokens comes from tenth sublist.

The following detail will be explained in the table and chart of the most frequent academic word.

Table 12 Table of the most dominant academic word in each sublist

Sublist	AWL	Tokens	Rank
1	Research	748	1
2	Text	230	2
3	Technique	156	3
4	Cycle	154	4
7	Grade	154	
9	Found	40	5
6	Interval	16	6
8	Drama	11	7
5	Equivalent	9	8
10	Odd	2	9

How is the Academic Word List (AWL) used in Undergraduate students' abstract journal article?

This research questions stands for give clear picture about how students' elaborate their writing especially in word choosing. Whether they are used academic word properly or not. For these analyses, researcher can get better explanation about the way students use the academic word. It is very important for data enrichment. To give best generalization whether students have problem in academic word submission in their abstract or not.

This research tells that from 3.500.000 academic words of text adopted from Coxhead (2000) Academic Word List (AWL), have been only 183 types of academic word used by them. It means only 0,00522%, students applied academic word in their abstract.

Which Academic Word sublist that is the most dominant used by Undergraduate students?

There are two (2) sublists which become the most dominant used by undergraduate students. They are *Sublist 1*and *Sublist 2* with 38 types of academic words. From the *Sublist 1* the academic words are;*research, data, method, process, role, factor, consist, actors, significant, approach, indicated, evidence, identification, indicate, analyzing, identify, consisting, context, specific, available, benefits, create, environment, function, identified, indicators, majors, roles, similar, source, area, economic, enviromental, occur, percentages, requirements, respond, and unstructed.*Then, from the *Sublist* 2 the academic words are *text, strategy, design, items, conducting, aspect, category, item, elements, positive, focus, participate, relevant, achieve, evaluating, journal, primary, acquired, assist, feature, transfer, achieving, acquire, affect, communities, consequently, culture, final, obtain, positively, potential, previous, range, ranging, site, sites,* and *survey.* The Most Frequent Academic Word in Each Sublist. Moreover, this is the table that shows variance of the amount of academic words for every sublist.

Table 13 Table of TheAmount of Academic Word used in each Sublist

Sublist	Amount of Academic Words used
1	38
2	38
3	17
4	24
5	19
6	15
7	11
8	13
9	6
10	2
Total	*183 academic words.*

How many non-academic word used by undergraduate students in their abstract journal article?

The amount of academic word is 183 word types from 10 sublist. It means only 0,0052% academic word used in students' abstract journal article. On the contrary, the amount of non-academic word is 1974 word types.

Discussion

In this section, the researcher tries to summarize the findings and explain more detail about how the data are analyzed using word list and keyword list in AntConc program. From this study, the researcher can argue that there are many type of part of speech tagset appear in students' abstract journal article. There are three types of the most frequent tagsets used by students. They are *noun, determiner,* and *preposition* tagset. The *noun* tagset has 18,2 %, *determiner* tagset has 12,8% and the last is *preposition* tagset with 8,2% as the whole percentage.

From this result, the most frequent tagset can be idintify, that is *NN* with 18,2% used in students' abstract journal article. By analyzing every single tagset exists in part of speech, the data results becomes more detail and accurate.

Every single tool has its own function and results. Asa table of specification, the researcher used 10 sublist promoted by AvrileCoxhead, to see the most frequent academic word appear in this research. Those are will be

attached in word list tool ad keyword list tool in the AntConc program, then it will automatically calculate based on the tagset code or word type which is run on the search button.

The sublist itself, give the important information about the frequent academic word using by students. From these analyses, the researcher can conclude the findings for this research. Moreover, researcher can give the clear picture about students' ability in writing scientific paper; the ability in using academic word. If the using of the academic word reach 50% of the sublist, it means the students' anxiety to string up the academic word in their writing is high. The other way, if the result is less than 50% of the sublist, it means the students' anxiety to string up the academic word in their paper is low. This case, can show students' lexical knowledge, whether they are familiar with the academic word then they can make good sentences in their abstract or not.

In fact that, the research shown only 0,0052% academic word used in studnets' abstract journal article. This condition leads the researcher to make final conclusion about the students' ability in used academic words in scientific wrtiing are still to be concerned. On the other side, the researcher will make final conclusion about the computation of part of speech. There are five concern while analyzed the part of speeh tagset; (1) the word types; (2) the word tokens; (3) the most frequent word; (4) the frequent word (fx >100); and the last is (5) the rare word (fx >50). With this formation, the results will be highly detail and structured.

From this result, with 51,4% as the component of part of speech in students' abstract jorunal articles shows that highly comprehensive in word distribution. As the evidences, the amount of 31 types of tagset (from 54 types of tagset exists before doing corpus analyzing) appeared in students' abstract journal article. Nonetheless, there is a very big concern for students to enhance their ability in making a better scientific writing. To do that, the students can lead themselves to have self-directed study using AntConc. Where AntConc can help them to answer their confusion relanvance to linguistics analysis. Moreover, AntConc can classified each tagset and academic word correctly by looking at the tagger or code.

Considering, the subject of research in this case are second language learner, then producing a scientific paper, for example abstract is big obstacle for them. Because, they need to take into consideration about the content of the abstract, the grammar and also word choice, lexicon, academic words. It is not easy. Choosing the right academic word can make the content of the abstract more interesting to be read.

CONCLUSION AND SUGGESTION

Conclusion

The analysis in finding session show that the there fifthy four (54) types of Tree Tagger Tagset appears in this research, covered thirty one (31) types for the parts of pseech tagset with the tokens which is the concern of this research, shows the important roles in scientific writing especially in student; abstract journal article. The results shows, *noun* as the highest frequency as the first place; *preposition*in the second place, *determiner*in the third place, *verb*in the fourth place, *adjective*in the fifth place, *conjunction*in the sixth place, *adverb* in the seventh place, and the last is *interjection*. The whole percentage is 51,5% in 208 students' abstract journal article; as the component of the part of speech in students' abtract journal article. The rest, 48,5% stands for the others tagset out of part of speech concern.

Meanwhile, to conclude the second research questions, the most dominant POS tagsetappeared during this research, is *noun* with 11.710 tokens or 18,2% percentage of *noun*. Moreover, to conclude the result of third research question, the most dominant word appear during this research is *research* with 748 tokens. Then, for the fourth research questions, from 3.500.000 academic words, have been only 183 types of academic word used by them. It means only 0,00522 %, students applied academic word in their abstract. In fifth research question, the result that

researcher got is Sublist 1 as the most dominant word sublist used by undergraduate students in their abstract journal article. And for the last research question, researcher need to compare the used of academic words and non-academic words using by undergraduate students. The comparison shows, the amount of academic words is 183 word types from 10 sublist. On the contrary, the amount of non-academic word is 1974 word types.

The researcher believes that with the following results show that students still faced difficulties in using academic words from the observation and analyzing that researcher did. Because of the lack of vocabularies especially in academic words, induce the poor quality of writing. In fact that, although writing subject is exist in the course, it doesn't mean the quality of students writing will be better than before. Researcher assumes that, with applying the use of AntConc as the language analyser program can help students to solve their problem in term of the lack of vocabularies especially academic words in this case. They can learn by themselves. Starting identifying every single academic words and the way to use them in the sentences. Then, by that activities, slowly but sure, they can master all academic words.

Because of the rapid changing of the technology, teachers, lectures even students can not denied learning with such of language analyser program becomes most helpful than learning traditionally or manually. It takes much time consuming, boring, and inaccurate. Then, it is a good idea to applying AntConc in learning a language.

Suggestion

The since this research can be useful to the development of teaching learning especially for analyze the language use in term of academic word, the researcher has some suggestions to all parties who will read, use, and improve this study. (1) To the next researcher who will conduct the same focus, that is analysing of the use of academic words, he / she should consider the sample sizes. With the huge sample sizes will produce multiple types of data. Do not afraid with the huge of data, because AntConc can solve it even with millions of data corpus. (2) To the next researcher who will continue this research, that is analysing the use of academic words, he / she can widen the coverage analysis into collocation statistics in term of MI (Mutual Information) score and *t*-score. (3) To the next researcher who will conduct the same focus, he / she can change the corpus sample into the complex one. For instance the students' thesis, or he / she can compare two corpus with different year of thesis collecting. (4) To the readers or users, especially the educators (teachers and lecturers), it is very important to more focus on such kind of this study. It is because the results of this research are very useful to enhance students' motivation and engagement in learning academic word list in order to create a good scientific writing. (5) To the readers or users, especially the educators (teachers and lecturers), it is important to start become a supervisor who promote the use of AntConc program in teaching learning context, because it is very helpful in language learning.

BIBLIOGRAPHY

Aston &Burnard. (1998). *The BNC Handbook: Exploring the British National Corpus with SARA*. Edinburgh: Edinburgh University Press

Biber, D. (2012). Corpus-based and corpus-driven analyses of language variation and use.*The Oxford Handbook of lInguistic Analysis,* 1. DOI: 10.1093/oxfordhb/9780199544004.013.0008

Dobric, N., Graf, E. &Onysko, A. (2016).*Corpora in applied linguistics: current approaches*. United Kingdom, UK: Cambridge Scholars Publishing

Ghomeshi, J., Paul, I. & Wiltschko, M. (2009). *Determiners: universals and variation*. Amsterdam: John Benjamins Publishing Company

Krug, M. &Schluter, J. (2013).*Research methods in language variation and change*. New York, NY: Cambridge University Press

McEnery, T., Xiao, R. &Tono, Y. (2006).*Corpus-based language studies: and advance resource book*. United States of America, USA:Routledge Applied Linguistics

Tognini-Bonelli, E. (2001). Corpus linguistics to work: studied in corpus linguistics. Amsterdam : John Benjamin Publishing Company

Valli, C. & Lucas, C. (2000). *Linguistics of American Sign Language : an introduction*. 3rd ed. Washington,D.C:USA: Clere Books, Gallaudet University Pr